DOCUMENT RESUME

ED 331 795 SP 033 014

AUTHOR Guskey, Thomas R.; Sparks, Dennis
TITLE Complexities in Evaluating the Effects of Staff Development Programs.
PUB DATE Apr 91
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 2-6, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Context Effect; Educational Environment; Elementary Secondary Education; Inservice Teacher Education; *Models; *Program Content; Program Effectiveness; *Program Evaluation; *Staff Development

ABSTRACT

Because of increased emphasis on accountability, program evaluations today must go beyond measures of change in program participants to consider the effects, either direct or indirect, of staff development on students and their learning. A model is presented illustrating the relationship between staff development for teachers and student learning outcomes, and the external factors that influence this relationship. Three factors are identified: (1) quality of the staff development program; (2) the content of the staff development program; and (3) the characteristics of the context in which the program is carried out. Although other models consider implementation as a separate factor, in this model quality or degree of implementation is considered a facet of the total process and therefore is a component of the first factor, the quality of the staff devlopment program. The potential effects of these factors on program evaluation results are described, along with procedures for estimating those effects. Finally, strategies are outlined, based on the model, for enhancing the quality and validity of staff development program evaluation. (IAH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED331795

Complexities in Evaluating the Effects of Staff Development Programs

Thomas R. Guskey & Dennis Sparks

Send correspondence to

Thomas R. Guskey
College of Education
University of Kentucky
Lexington, KY 40506
Phone: 606/257-8666

BEST COPY AVAILABLE

Paper present at the annual meeting of the
American Educational Research Association,
Chicago, IL, April, 1991

SP033014

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Thomas R Guskey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Complexities in Evaluating the Effects
of Staff Development Programs

Although modern proposals for educational reform vary widely in their scope and content, nearly all emphasize the need for high quality staff development. Regardless of the way schools are structured or restructured, staff development will be essential to the improvement process. Educators at all levels need to keep abreast of the new knowledge in their field, especially since that knowledge is expanding today at an ever increasing rate. Furthermore, they need to upgrade their professional skills on a regular and ongoing basis so that they can implement that new knowledge in effective and efficient ways.

Along with the emphasis on high quality staff development, the current wave of reforms also stresses the need for greater accountability in education. With regard to staff development, this press for greater accountability is most evident in program evaluation procedures. No longer is it considered adequate to implement a large-scale staff development program and then simply document what was done (e.g., Seventy percent of faculty members took part in a series of workshops on classroom management skills). It is also considered insufficient to evaluate staff development programs only in terms of their effects on the

educators who took part (e.g., As a result of the program, 70 percent of faculty members reported reduced levels of stress).

Demands for accountability require that staff development program evaluations focus instead on the programs' impact on students, and especially the results yielded in terms of improved student learning outcomes. Any valid improvement effort should, after all, benefit the constituency our educational system is principally designed to serve. Therefore, the bottom line in the evaluation of any staff development program or policy ought to be, "What will this mean for students and how will it benefit them?"

But extending evaluations of staff development programs to consider impact on student learning is not a simple task. The relationship between staff development and improvement in student outcomes is far more complicated than is generally assumed. Factors external to the staff development process can impinge on this relationship. While the influence of these factors can be great, it is typically ignored in staff development program evaluations as well as in research studies of the staff development process.

Rationale for a New Model

For most staff developers, the demand for greater accountability and the accompanying emphasis on learner outcomes

will require a significant change in the way programs are evaluated. Although evaluations of staff development programs have long been criticized as being short-sighted (Howey & Vaughan, 1983), even those regarded as exemplary usually take into account only effects on participating teachers (Showers, Joyce, & Bennett, 1987). Outcome measures gathered in program evaluations are generally restricted to indices of change in the way these educators think, what they believe, and what they do as a result of their participation in the program.

While it may be valuable and necessary to document changes such as these, accountability demands make it clearly not enough. Program evaluations today must go beyond measures of change in program participants to consider the effects of staff development, either direct or indirect, on students and their learning. Efforts must be made to determine whether or not staff development programs result in meaningful improvement in how well students learn, in the way they learn, or in how they feel about themselves as learners.

Presented in this paper is a model illustrating the relationship between staff development for teachers and student learning outcomes. Also illustrated in the model are the external factors that influence this relationship. The potential effects of these factors on program evaluation results are described, along with procedures for estimating those effects. Finally, strategies are outlined, based on the model, for

enhancing the quality and validity of staff development program evaluations.

The Model

Studies conducted over the last two decades have offered many valuable insights into the aspects of staff development programs that contribute to desired change in the behaviors and instructional practices of teachers (Doyle & Ponder, 1977; Gall & Renchler, 1985; Guskey, 1986; Huberman & Miles, 1984; Joyce & Showers, 1988). Nevertheless, relative few investigations have extended this line of inquiry to determine whether these changes in teachers' behaviors and practices do, in fact, result in desired improvements in student learning.

But evaluating the impact of staff development on learning outcomes cannot be accomplished, as some may think, simply by adding pre- and post-measures of student learning to evaluation designs. The complex nature of the relationship between staff development and improvements in student outcomes confound such measures. A variety of factors influence this relationship. Some of these are unique to the setting and undoubtedly lie outside the control of those planning or implementing the staff development program. A school district's calendar or personnel policies, for example, might restrict what can be done. Still, other factors known to be highly influential, like the particular training procedures employed, are within the control of staff

developers, directly alterable, and need to be considered when evaluating the results of staff development efforts.

Illustrated in Figure 1 is a model describing factors that impinge on the relationship of staff development and student learning outcomes. As the figure shows, the quality of the staff development program itself has a direct and primary influence on the improvement of student outcomes. As the quality of staff development programs is enhanced, resulting improvements in student learning are likely to be greater.

Insert Figure 1

In addition to program quality, the content of the staff development program and the characteristics of the context in which the program is carried out also can be highly influential. The effects of these two factors can be direct, interactive, or both. Furthermore, although their effects can often be measured and, under some conditions, accounted for or controlled, it seems unlikely their influence can ever be totally eliminated.

Upon first inspection, this model may seem overly simplistic. Yet its simplicity is not meant to impugn the complexity of the relationship between staff development and improvement in student learning. It may be that the factors

included in the model do not capture all the elements that influence this relationship, and other important factors may exist. The model should not be taken, therefore, as totally comprehensive. It is offered principally as a working framework from which to understand better this complex relationship, to guide future investigations and, hopefully, to improve the quality and validity of staff development program evaluations.

Quality of the Staff Development Program

Obviously, the quality of the staff development program will have a strong and direct influence on any improvements that result in student learning. Program quality is also the factor most directly alterable by staff developers. Although research on the exact nature of the influence of program quality on student learning is not extensive, investigations on program implementation offer some general notions about elements that are likely to be important.

In their early work on teacher decision-making, for example, Doyle and Ponder (1977) suggested that the manner in which an innovation is presented to teachers affects their implementation decisions. Three criteria were believed to be particularly important. The first they labeled instrumentality, which refers to how clearly and specifically the practices are presented. The second they suggested was congruence, which describes how well the new practices are aligned with teachers' present teaching

philosophy and practices. The third they believed was the cost, which they defined as teachers' estimate of the extra time and effort the new practices require compared to the benefits such practices are likely to yield. Later studies by Mann (1978) and Mohlman, Coladarci, and Gage (1982) generally confirmed the importance of these elements and showed how they can be used to enhance the quality of staff development programs.

More recent studies by Bennett (1987) and Joyce and Showers (1983, 1988) have identified additional components that appear to be shared by staff development programs that result in classroom implementation. These components include the presentation of theory, modeling or demonstration, practice under simulated conditions, structured and open-ended feedback, and coaching for application. Although the relative importance of some of these components has been questioned in other investigations (Sparks, 1983; Sparks & Bruder, 1987), it is evident that consideration of these elements is likely to enhance the quality of any staff development program and, as a result, lead to greater improvements in student learning.

Some might argue that the quality of program implementation is separate factor that should be taken into account when considering the relationship between staff development and improvement in learning outcomes. Indeed, many staff development program evaluations include measures of "degree of implementation" to verify that the new ideas or techniques were

actually incorporated in classroom practice. In the model presented here, however, staff development is considered to be a multifaceted process. As such, it is envisioned to include not only initial training, but also the readiness activities that precede training, the practice and coaching that take place during training, as well as the follow-up and support activities that take place during program implementation. Therefore, quality or degree of implementation is considered one facet of this process and, thus, a component of the quality of the staff development program.

Program Content

Another major factor shown in the model to influence the relationship between staff development and student learning outcomes is the content of the staff development program. More specifically, it is the effectiveness of the particular set of ideas or the particular innovative strategy upon which the staff development activities focus. Not all innovations are created equal. Some have a very extensive research base while others have virtually none. Of those that do, some have been found to have a very powerful impact on student learning while others appear to have relatively modest effects (see Bloom, 1984; Fraser, Walberg, Welch, & Hattie, 1987; Walberg, 1984a, 1984b, 1990).

The magnitude of an innovation's effect on student outcomes is often estimated today through a technique called "meta-analysis" (Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985). In conducting a meta-analysis, researchers first gather all of the high quality studies of an innovation that are available. For each study the results attained by students who took part in the innovation (the experimental group) are compared to those of students who did not (the control group). The standardized difference in results between these two groups of students is referred to as the "effect size." Thus if the experimental did much better on a particular outcome than did the control group, the effect size would be large. If, on the other hand, the difference between the two groups is relatively modest, the effect size would be small. By calculating the average effect size across all the high quality studies collected, researchers are able to come up with an estimate of the typical effect size for that innovation on specific student outcomes. Assuming that this average effect size is calculated through procedures that are unbiased and reliable, it can then be used to compare the relative impact of different innovations.1

But when researchers conduct a meta-analysis, synthesizing the results from many studies to determine the average effect

1 It has been noted that certain procedures used to calculate effect sizes, particularly Slavin's (1986) "best-evidence" synthesis methods, may yield estimated effect sizes that are neither unbiased nor reliable (Guskey, 1987; Heibert, 1987; Joyce, 1987; Kulik, Kulik, & Bangert-Drowns, 1990b; Walberg, 1988).

size of a particular innovation, they generally ignore the quality of the staff development that was involved. Most make the assumption, either explicitly or implicitly, that the quality of the training used to introduce the innovation and the nature of the follow-up support provided to educators as they implemented the new ideas, had either no effect on student learning or an effect that was constant across all studies. Although this allows the effect size of an innovation to be estimated with great precision, it disregards what is likely to be a very powerful intervening influence.2

Researchers investigating factors that contribute to the quality of staff development programs, on the other hand, generally focus on training components that are common across programs of widely varied content. These researchers are primarily concerned with the characteristics of the training and follow-up activities that lead to implementation, regardless of the particular set of ideas or the innovation involved. In their efforts to identify factors that are generalizable to a broad range of staff development endeavors, they combine results from programs dealing with a variety of innovations, ignoring differences in the relative effectiveness of those innovations.

---

2 It should be noted that the accuracy of an estimated effect size is also dependent upon the quality of the research designs used in the selected studies, the size of the sample, and the reliability of the measures of student learning employed.

Both these approaches appear to have shortcomings when considering the nature of the relationship between staff development and student learning outcomes. Differences in the quality of staff development leading to the implementation of a particular innovation may contribute to inconsistency in the calculation of an effect size for that innovation. This may, in fact, be one reason why effect size estimates for the same innovation often vary greatly from study to study (Hedges & Olkin, 1985). Similarly, failure to consider the effectiveness of the particular innovation that is the topic of a staff development program may lead to erroneous conclusions about the effectiveness of particular training and follow-up activities. In other words, the staff development program might have been conducted very well but led to no improvement in student learning because the innovation upon which the training focused was ineffective.

Context Characteristics

A third factor described in the model that is believed to influence the relationship between staff development and student learning is the context in which the program is conducted and implementation takes place. Extensive research evidence on program implementation shows that organizational culture and climate can strongly influence both initial implementation and the continued use of any set of new ideas or innovative strategies (Joyce, 1990). In a large-scale study of federally

sponsored programs, for example, Berman and McLaughlin (1978) found that successful programs generally took place in environments characterized by strong administrative support for teachers coming from both principals and superintendents (see also McLaughlin, 1990). Similarly, Little's (1981) study on the effects of staff development showed that programs were most likely to be successful where there was "a norm of collegiality and experimentation." Contexts that nurture support and trust, encourage shared decision-making and responsibility, and provide ongoing assistance and problem solving appear best in sustaining successful improvement efforts (Little, 1982).

Although contextual characteristics such as these are known to be influential, they, too, are generally ignored in research studies on staff development as well as in evaluations of staff development programs (Fullan, 1990). Again, because staff development researchers are typically interested in identifying the characteristics of successful programs that are generalizable to a variety of settings, any detailed consideration of context differences is often passed over. Likewise, context characteristics are seldom considered in evaluations of staff development programs. Those evaluations that take the form of indepth case studies are, of course, rare exceptions to this general rule. But while case studies are a source of rich information, they are also frustrating to those interested in implications that are more broadly applicable.

Estimating Effects

Recognizing the influence of these factors and their possible confounding effects is one thing. Estimating the precise magnitude of their influence or controlling for it is quite another. Although doing so is possible, it can require skills and additional resources far beyond those available to most staff developers. In addition, the procedures necessary to estimate the effects of these factors often introduce artificial constraints in an evaluation design. As a result, what is gained in evaluation precision may be lost in diminished validity and utility of the findings.

One way to control or account for the influence of program content, for instance, would be to restrict staff development training activities to only those ideas or innovations for which substantial research evidence has been compiled and synthesized. In this way, the magnitude of effect on student outcomes achieved by that innovation through a staff development program of "average" quality could be anticipated, based on the results from previous studies. Excellent summaries of the innovative strategies that have been so thoroughly investigated are offered by Bloom (1984) and Walberg (1984b). This procedure would not only provide a means to distinguish the influence of content from staff development program quality, it is also likely to enhance the prospects for program success.

It is important to recognize, however, that making this restriction would greatly reduce the number of program content options available to staff developers. Although many of today's educational innovations are described as "research-based," this does not mean their impact on student learning outcomes has been thoroughly investigated. In fact, relative few of the innovative strategies that are currently in vogue and the focus of many staff development programs have been extensively or systematically studied. Two notable exceptions are cooperative learning (Johnson & Johnson, 1989) and mastery learning (Guskey & Pigott, 1988; Kulik, Kulik, & Bangert-Drowns, 1990a). In most cases when an innovation is described as "research-based," it simply means the creators of that innovation referred to some body of research literature when initially formulating their ideas. Those innovative strategies that have their own research base; that is, that have been carefully implemented in a variety of settings and their impact on student outcomes systematically evaluated, are far fewer in number.

Another way to separate the effects of program content in evaluations of the quality of staff development programs would be to hold the content constant while varying aspects of the selection, training, and follow-up activities. This technique is sometimes referred to as "planned variation." It would be accomplished by taking a well defined program, one specific model of cooperative learning, for instance, and systematically altering the staff development activities used to introduce the

model and support its implementation. Since the program content remains the same, any variation in the improvement in student learning outcomes that result could be attributed to differences in the quality of the staff development.

Confounding both of these approaches, however, is the possible interactive influence of context characteristics. One might argue, for example, that organizational culture and climate are likely to influence the appropriateness and, hence, effectiveness of any innovation, despite the research evidence supporting it. Likewise, the success of particular staff development activities might vary greatly depending upon differences in relationships between administrators and faculty, the type or size of the school, or the kinds of students served.

Complications such as these might cause some to throw up their hands and give up on the process of evaluation all together. After all, with so many confounding factors, how can the results from any staff development program evaluation be considered truly valid or reliable. But while it is very complex, the situation is not hopeless. It is, however, somewhat analogous to the "uncertainty principle" in physics.

According to the uncertainty principle, developed by German physicist and 1932 Nobel Laureate Werner Heisenberg, either the position or the momentum of a subatomic particle can be measured with accuracy, but the accuracy with which both can be measured

simultaneously is limited. In other words, the more accurately one determines the position of the particle, the less one knows about its momentum. Conversely, the more accurately one determines its momentum, the less one knows about the particle's position. Thus while physicists take great pride in the exactitude of their science, they find some uncertainty is absolutely necessary, due to the nature of the phenomenon they study and limitations in their measurement devices.

Similarly, those who evaluate staff development programs also must accept some amount of uncertainty. Determining the exact nature of the influence of program content and context characteristics is likely to be impractical in many instances and impossible in others. Still, this limitation should not deter staff development program evaluators from recognizing the potentially powerful influence of these factors, documenting or measuring their influence whenever possible, and considering their impact when interpreting evaluation results to all interested parties.

Implications for Improving the Quality of Program Evaluations

Staff development program evaluation is obviously more complex than it may appear at first glance, especially if the purpose of the program is to produce significant, lasting improvements in student learning outcomes. While in the past it may have been sufficient for program evaluators to focus on

assessing change only in the attitudes, knowledge, or behaviors of educators, today the discipline of staff development has reached a degree of sophistication that requires a far more complex approach to program evaluation.

It is critically important for evaluators today to collaborate closely with program planners and practitioners from a program's inception. As a part of this collaboration, evaluators should help focus attention on questions that not only will be helpful in the collection of meaningful evaluation data, but also will assist in the development of programs of sufficient magnitude and power to affect student outcomes. These questions might include:

Is the staff development program driven by clearly stated, measurable district or school objectives?

Is a systemic view of the change process expressed in the program's plans? That is, is it recognized that change in one part of the system affects all other parts?

Are all appropriate parts of the organization contributing to the change effort? For example, is there parent involvement? Curriculum revision? Changes in supervisory practices?

Is the staff development program's content sufficiently grounded in research to ensure that if properly implemented it will produce the desired changes in student outcomes?

Thoughtful consideration of questions such as these when staff development programs are being planned will increase the likelihood that these programs, faithfully implemented, will produce the intended results.

Evaluations of staff development programs can greatly improve program quality and, as a result, make a lasting contribution to the field of staff development. To do so, however, program evaluators will need to focus on student learning outcomes and recognize that success will require rigorous attention to program content and systemic contextual factors as well as to staff development processes.

References

Bennett, B. B. (1987). The effectiveness of staff development training practices: A meta-analysis. Unpublished doctoral dissertation, University of Oregon.

Berman, P., & McLaughlin, M. W. (1978). Federal programs supporting educational change. Vol. VIII: Implementing and sustaining innovations. Santa Monica, CA: Rand Corporation.

Bloom, B. S. (1984). The search for methods of group instruction as effective as one-to-one tutoring. Educational Leadership, 41(8), 4-18.

Coladarci, T., & Gage, N. L. (1984). Effects of a minimal intervention on teacher behavior and student achievement. American Educational Research Journal, 21, 539-555.

Doyle, W. & Ponder, G. (1977). The practical ethic and teacher decision-making. Interchange, 8(3), 1-12.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. International Journal of Educational Research, 11, 73-145.

Fullan, M. G. (1990). Staff development, innovation, and institutional development. In B. Joyce (Ed.), Changing school culture through staff development. 1990 yearbook of the Association for Supervision and Curriculum Development. Alexandria, VA: Association for Supervision and Curriculum Development.

Gall, M. D., & Renchler, R. S. (1985). Effective staff development for teachers: A research based model. Eugene, OR: ERIC Clearinghouse on Educational Management, University of Oregon.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Guskey, T. R. (1986). Staff development and the process of teacher change. Educational Researcher, 15(5), 5-12.

Guskey, T. R. (1987). Rethinking mastery learning reconsidered. Review of Educational Research, 57, 225-229.

Guskey, T. R., & Pigott, T. D. (1988). Research of group-based mastery learning programs: A meta-analysis. Journal of Educational Research, 84, 197-216.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

Heibert, E. H. (1987). The context of instruction and student learning: An examination of Slavin's assumptions. Review of Educational Research, 57, 337-340.

Howey, K. R., & Vaughan, J. C. (1983). Current patterns of staff development. In G. A. Griffin (Ed.), Staff development. Eighty-second yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Huberman, M., & Miles, M. B. (1984). Innovation up close: How school improvement works. New York: Plenum.

Johnson, D. W., & Johnson, R. T. (1989). Cooperation and competition: Theory and research. Edina, MN: Interaction Press.

Joyce, B. (1987). A rigorous yet delicate touch: A response to Slavin's proposal for 'best-evidence' reviews. Educational Researcher, 16(4), 12-14.

Joyce, B. (Ed.) (1990). Changing school culture through staff development. 1990 yearbook of the Association for Supervision and Curriculum Development. Alexandria, VA: Association for Supervision and Curriculum Development.

Joyce, B., & Showers, B. (1983). Power in staff development through research on training. Arlington, VA: Association for Supervision and Curriculum Development.

Joyce, R. B., & Showers, B. (1988). Student achievement through staff development. New York: Longman.

Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990a). Effectiveness of mastery learning programs: A meta-analysis. Review of Educational Research, 60, 265-299.

Kulik, J. A., Kulik, C. C., & Bangert-Drowns, R. L. (1990b). Is there better evidence on mastery learning? A response to Slavin. Review of Educational Research, 60, 303-307.

Little, J. W. (1981). School success and staff development: the role of staff development in urban desegregated schools. Executive summary. Washington, DC: National Institute of Education.

Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. American Educational Research Journal, 19, 325-340.

Mann, D. (1978). The politics of training teachers in schools. In D. Mann (Ed.), Making change happen. New York: Teachers College Press.

McLaughlin, M. W. (1990). The Rand change agent study revisited: Macro perspectives and micro realities. Educational Researcher, 19(9), 11-16.

Mohlman, G., Coladarci, T., & Gage, N. (1982). Comprehension and attitude as predictors of implementation of teacher training. Journal of Teacher Education, 33(1), 31-36.

Showers, B., Joyce, B., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. Educational Leadership, 45(3), 77-87.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analysis and traditional reviews. Educational Researcher, 15(9), 5-11.

Sparks, G. M. (1983). Synthesis of research on staff development for effective teaching. Educational Leadership, 41(3), 65-72.

Sparks, G. M., & Bruder, S. (1987). Before and after peer coaching. Educational Leadership, 45(3), 54-57.

Walberg, H. J. (1984a). Synthesis of research on teaching. In M. C. Wittrock (Ed.), Handbook of research on teaching. Washington, DC: American Educational Research Association, pp. 214-229.

Walberg, H. J. (1984b). Improving the productivity of America's schools. Educational Leadership, 41(8), 19-27.

Walberg, H. J. (1988). Response to Slavin: What's the best evidence? Educational Leadership, 46(2), 28.

Walberg, H. J. (1990). Productive teaching and instruction: Assessing the knowledge base. Phi Delta Kappan, 71, 470-478.

Factors Affecting the Improvement of Student Learning Outcomes Through Staff Development